DOCUMENT RESUME

ED 352 042 IR 054 264

AUTHOR Waters, Donald; Weaver, Shari
TITLE The Organizational Phase of Project Open Book.
INSTITUTION Yale Univ., New Haven, Conn. Library.
SPONS AGENCY Commission on Preservation and Access, Washington, DC.
PUB DATE Sep 92
NOTE 14p.
PUB TYPE Reports - Descriptive (141)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Access to Information; Books; College Libraries; Computer Printers; Higher Education; *Information Technology; *Microfilm; Planning; *Preservation; *Research Libraries
IDENTIFIERS *Digital Imagery; *Electronic Text; Vendors; Xerox Corporation; Yale University CT

ABSTRACT

The Yale University Library is now organized to move ahead with Project Open Book, the conversion of 10,000 books from microfilm to digital imagery. In the first phase of the Project--the organizational phase--a Steering Committee was established that included several faculty members, and a project team was created. In addition, Yale conducted a formal bid process and selected the Xerox Corporation to serve as its principal partner in the project. Xerox has identified for Yale the required equipment, software, and services to complete the project, as well as their costs, and has proposed a three-phase implementation plan. The implementation will ultimately result in a conversion subsystem, browsing stations distributed on the campus network within the Yale Library, a subsystem for storing 10,000 books in digital form, and network access to high-quality image printers. The process leading to the selection of the vendor helped isolate areas of risk and uncertainty as well as key issues to be addressed during the life of the project. The Yale Library is now prepared to select the material for conversion to digital image form and to seek funding, initially for the first phase, and then for the entire project. This report reviews the purpose and scope of the project, outlines the steps taken in the first (organization) phase, and presents a summary of the results to date. (Author/KRN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED352042

The Commission on Preservation and Access

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

THE ORGANIZATIONAL PHASE OF PROJECT OPEN BOOK

A report to the Commission on Preservation and Access

by

Donald Waters
Director
and
Shari Weaver
Assistant to the Director
Library and Administrative Systems
Yale University Library

September 1992

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Maxine Sitts

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

11054264

1400 16th Street, N.W., Suite 740, Washington, D.C. 20036-2217 • (202) 939-3400

A private, nonprofit organization acting on behalf of the nation's libraries, archives, and universities to develop and encourage collaborative strategies for preserving and providing access to the accumulated human record.

BEST COPY AVAILABLE

Published by
The Commission on Preservation and Access
1400 16th Street, NW, Suite 740
Washington, DC 20036-2117

September 1992

Reports issued by the Commission on Preservation and Access are intended to stimulate thought and discussion. They do not necessarily reflect the views of Commission members.

Additional copies are available from the above address for $5.00. Orders must be prepaid, with checks made payable to "The Commission on Preservation and Access," with payment in U.S. funds.

This paper has been submitted to the ERIC Clearinghouse on Information Resources.

The paper in this puolication meets the minimum requirements of the American National Standard for Information Sciences-Permanence of Paper for Printed Library Materials ANSI Z39.48-1984.

COPYRIGHT 1992 by the Commission on Preservation and Access. No part of this publication may be reproduced or transcribed in any form without permission of the publisher. Requests for reproduction for noncommercial purposes, including educational advancement, private study, or research will be granted. Full credit must be given to the author(s) and The Commission on Preservation and Access.

The Organizational Phase of Project Open Book

Introduction

The Yale University Library is now organized to move ahead with Project Open Book, the conversion of 10,000 books from microfilm to digital imagery. In the first phase of the Project—the organizational phase—Yale established a Steering Committee, including several faculty members, and created a project team. In addition, Yale conducted a formal bid process and selected the Xerox Corporation to serve as its principal partner in the project. Xerox has identified for Yale the required equipment, software and services to complete the project, as well as their costs, and has proposed a three-phase implementation plan. The implementation will ultimately result in a conversion subsystem, browsing stations distributed on the campus network within the Yale Library, a subsystem for storing 10,000 books in digital form, and network access to high-quality image printers. The process leading to the selection of the vendor helped isolate areas of risk and uncertainty as well as key issues to be addressed during the life of the project. The Yale Library is now prepared to select the material for conversion to digital image form and to seek funding, initially for the first phase and then for the entire project.

In this report we review the purpose and scope of the project, outline the steps taken in this first, organizational phase and, finally, present a summary of the results to date.

Project purpose and scope

The June 1991 report entitled *From Microfilm to Digital Imagery* constitutes the master plan for Project Open Book, a major effort in the Yale University Library to explore the usefulness of digital technologies for preserving and improving access to deteriorating documents.[1] The planned project is founded on a vision of the research library of the future as an institution whose mission is to generate, preserve and improve for its clients ready *access*—both intellectual and physical—to recorded knowledge. The place of electronic tools in the library of the future will depend on how well they measure up against this mission. Among the various types of electronic information that might find a place in the access-oriented library of the future are documents in digital image form. Yale is motivated in Project Open Book to test and explore a set of hypotheses about the feasibility of digital imaging as a preservation tool. These working hypotheses are based, ultimately, on an "ideal" model of digital image documents in the library of the future.

The purpose of Yale's Project Open Book is related in many ways to Cornell's CLASS Project, but Open Book is distinguished from CLASS in terms of its scale and the source of its documents. In Project Open Book, Yale will seek to convert 10,000 volumes into digital image form. This number is an order of magnitude larger than the number of documents originally converted at Cornell and our intention is to explore the effects of scale on emerging preservation imaging systems. In addition, rather than scanning directly from the original paper document, the purpose of the Yale project is to convert material from microfilm to digital image form and thus explore the promise that once we have preserved materials on film we can eventually and satisfactorily convert those documents into digital form. Several other working hypotheses also help to define and limit the scope of Project Open Book. Among the other key hypotheses are these:

- Microfilm is satisfactory as a long-term medium for preserving content;
- Digital imagery can improve access to recorded knowledge through printing and network distribution at a modest incremental cost over microfilm;
- Researchers will demand greater access to digital image libraries that contain thematically-related materials;
- Capturing and storing documents in digital image form is a necessary step leading to even further improvements in access (e.g., through the application of OCR).

Working hypotheses

The first working hypothesis—that microfilm is satisfactory as a long-term medium for preserving content—builds on the features of microfilm as a long-lasting, inexpensive technology that is well understood in libraries. However, the linear nature of microfilm does not provide easy access. It is cumbersome to browse and read, it requires special equipment at a single location, it does not facilitate use of an item's internal structure, and it does not produce high quality paper copies.

The deficiencies of microfilm lead to the second working hypothesis—that digital imagery can improve access to recorded knowledge through printing and network distribution at a modest incremental cost over microfilm. In theory, users of existing and widely distributed workstations could gain access to a digital image database from outside the library and, indeed, from around the world over existing high-speed networks. As resources allow, users could employ, again in theory, relatively simple commands to gain ready and random access to documents in digital image form by actual and relative page numbers and to internal document structures such as the table of contents, notes, chapters and indices. Moreover, the quality of the printed output from an image database is likely to be indistinguishable by the human eye from a printed book.

If documents in digital image form do prove more accessible than those in microform, its seems plausible to posit a third hypothesis—that researchers will demand greater access to digital image libraries containing thematically-related materials. In the conversion of documents to image form, we may impair physical access by disturbing the existing collocation schemes in libraries. A large collection of thematically-related materials in digital image form may help overcome the potential disadvantage in imaging of creating yet another source for students and scholars to look for relevant materials. In any case, investigation of this hypothesis will help insure that image access products and services integrate well into the daily routines of scholarly work and that they meet the performance and other delivery requirements of the user community.

Ultimately, technical processes such as optical character recognition (OCR), which convert digital images into full text, have the potential of greatly expanding intellectual access to documents in digital image form. Character recognition processes are not yet practical but investigation of the fourth working hypothesis—that capturing and storing documents in digital image form is a necessary step leading to even further improvements in access—will insure that Project Open Book does not preclude foreseeable future enhancements, such as OCR, and may help pave the way for their eventual application.[2]

An 'ideal' model of digital imagery

Taken together and verified in efforts like Project Open Book, these various hypotheses may lead to the specific conclusion that research libraries will choose, given a mix of flexible technology, to maintain information on microfilm for long-term preservation and in digital image form for ease of access. Otherwise, the purpose and scope of Project Open Book is designed to

lead research libraries like Yale's closer to a more general ideal model, originally outlined in *From Microfilm to Digital Imagery*.[3] The model posits the existence of an image document library that is created from multiple sources and with multiple uses (see Figure 1).

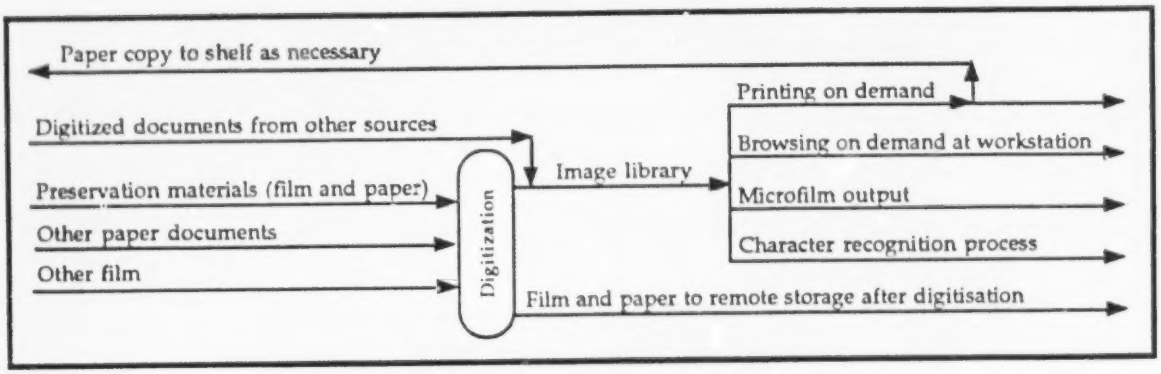

Figure 1. Digital Imagery in the Library

Eventually, we imagine, the library will itself generate digital image documents from film and paper for preservation purposes as well as for more general reasons, such as the creation of reserve materials or customized books of course readings. The library may also acquire image documents from external sources, such as service bureaus hired to reformat preservation materials or directly from publishers or vendors. After digitization, the library may opt to move the film and paper to remote storage. Users may then print documents from the image library, browse them at a workstation, or reformat them, say, by generating microfilm or by submitting them to a character recognition process. The quality—measured primarily in terms of resolution—of the image documents that the library generates and maintains will depend in large part on the expected mix of these various uses in both the long and short term.

Project Open Book, Phase One: Organization

In the first, organizational phase of Project Open Book, the objectives we set for ourselves included the following:

- create a project team and a steering committee, including faculty;
- find ways to test the validity of our original master plan;
- identify the specific hardware, software, services and staff that the project will require;
- establish the costs of the project and create a budget;
- select a vendor partner or partners;
- identify criteria for selecting the materials to convert from microfilm; and
- raise the necessary funds to support the next and subsequent phases of the project.

The Steering Committee and Project Team

The Steering Committee has met formally twice, once in November 1991 and again in May 1992. Millicent Abell, the University Librarian, chairs the committee. The Associate University Librarians, Michael Keller, Gerald Lowell and Karin Trainer, sit on the committee. Other members from the Library include Donald Waters, the project manager and Director of Library and Administrative Systems in the library, and the Head of the library's Preservation

Department (Marcia Watt before June 1, 1992; Paul Conway on and after that date). Richard Ferguson, the University Director of Computing and Information Systems, and Philip Long, the Director for Academic Computing in Yale's Computing and Information Systems Department, also serve on the Steering Committee. The faculty members are Nancy Cott from the History Department, Edward Tufte from the Political Science Department and Graphic Design, William Nordhaus from the Economics Department, and Paul Bracken from the School of Organization and Management.

The project team consists of Donald Waters, the project manager, the Head of Preservation (Marcia Watt before June 1, 1992; Paul Conway on and after that date), Philip Long, and Shari Weaver, the Assistant to the Director of Library and Administrative Systems. We originally intended our project team, with the help and advice of the steering committee, to perform many of the tasks identified for the first phase of Project Open Book. To help defray the costs of staffing the project team, the Library applied for and received $50,000 in funds from the Commission on Preservation and Access. As the team focused on selecting a vendor, however, it realized that it could use the Commission funds, in effect, to hire competing vendors each to generate a formal requirements analysis for Project Open Book and thereby to confirm the validity of Yale's original master plan. Given the formal requirements analysis as a basis, each vendor then could generate a competitive proposal for participating in the project and, in the process, also perform much of the other organizing work, such as identifying the hardware, software and services needed for the project and establishing the project costs.

Sequence of organizational work

The sequence of work during this organizational phase of the project began in July 1991, shortly after the publication of *From Microfilm to Digital Imagery*, when the Yale Library issued a formal request for proposal. The request identified two categories of response: one from vendors who wanted to contribute resources to the project and participate as partners; the other from vendors who wanted to bid imaging system equipment or software for sale to Yale. Vendors could, of course, respond in either or both categories.

The vendors responding to the request had to demonstrate, among other things, their fiscal and organizational stability. They also had to show their commitment to image document management involving a variety of document types. This criterion eventually disqualified many vendors who are in the business primarily to manage office documents. Office document management typically operates on the assumption that documents are organized with several pages in a folder and that folder is in turn stored in a file drawer in a file cabinet. Although many archival materials can, or could be made to, conform to such a metaphor, books, serials, pamphlets and other library documents simply do not so conform.

Another requirement for responding vendors was that they had to indicate their ability to serve as a systems integrator in a large-scale project over a multi-year period. They had to demonstrate an interest in developing any requisite software as a standard product for a general market. And they had to be willing to invest substantial resources of their own in the project. The nature of these criteria led us primarily, though not exclusively, to the large computer products vendors. The vendors from whom the Yale Library solicited proposals included IBM, Digital Equipment Corporation, Xerox, Eastman Kodak, Wang Laboratories, Apple Computer, NeXT Computer, Sun Microsystems, Minolta, Fuji, Mekel Engineering, W.J. Schafer Associates, Micro Dynamics, Oracle Corporation,

NOTIS Systems, Storage Technology Corporation, West Coast Information Systems, Accessible Archives, the Spaulding Company, and McGraw Hill, Inc.

While waiting for the vendor responses, which were due in September 1991, the Yale Library prepared and submitted its funding proposal to the Commission on Preservation and Access. By October, IBM, Digital Equipment Corporation, which coined the name Project Open Book, Xerox Corporation, Wang Laboratories and Associated Microfilm Services, a Fuji vendor, had all responded to Yale's request for proposal. Of these five, only Digital, Xerox and Associated Microfilm Services offered a detailed plan for establishing a partnership relationship with the Yale Library for the development of the imaging project. The project team tried to arrange qualifying visits to the offices of each of these three vendors. During October 1991, Donald Waters and Philip Long visited the Virginia offices of Xerox; during November they visited the Massachusetts offices of Digital Equipment. They were not able to meet representatives from Associated Microfilm Services, however, and in October the firm was dropped from further consideration as a development partner. Digital Equipment and Xerox thus remained as competing vendors with substantial and well-qualified proposals.

By the end of October, after the Commission on Preservation and Access agreed to fund this phase of the imaging project, the strategy for dealing with Digital Equipment and Xerox crystallized. As a condition of its partnership, each vendor had asked the Yale Library to pay for a detailed analysis of the requirements for Project Open Book. After consulting the Steering Committee in November and with support of the Commission funds, the Library agreed to pay both vendors for the requirements analysis on the further condition that each vendor submit, as a result of its work, a competitive proposal to complete the project. On the merits of the proposals, Yale would select one of the vendors as a principal project partner. Each vendor agreed that the one selected would rebate to Yale half of the costs of the initial requirements analysis.

By early February 1992, the Yale Library had negotiated and signed contracts with each vendor. According to the contracts, each company agreed that its proposal to Yale would meet four minimal requirements by providing:

1. a detailed specification of the required configuration of hardware, software, staff and services for the follow-on-project;
2. a phased implementation schedule;
3. a statement of responsibility for all phases for both Yale and the company;
4. and a detailed estimate of total cost per phase and overall project cost.

By early May, both vendors had completed their analyses and submitted their proposals to the Yale Library. By the end of May, the project team had evaluated the proposals, consulted members of the Project Steering Committee, and selected Xerox as the vendor of choice and partner in development for the Yale Library's digital imaging system.

Summary of Results

In this first, organizational phase of Project Open Book, two objectives remain to be achieved, namely, to identify criteria for selecting the materials to convert from microfilm and to raise the necessary funds to support the next and subsequent phases of the project. Otherwise, the Steering Committee and Project Team are in place and the Library has selected a vendor partner. In

addition, through the vendor selection process, Yale has achieved several other critical objectives.

One result of the selection process is proving to be very important for the project, but was not specifically set as an objective and, in retrospect, should have been. Our original plan focused on the development of a small team of experts in the project team. However, the process of working with the vendors to generate their analyses of our requirements forced us to realize that the success of the project depends on the informed participation of library and computer professionals with expertise ranging from technical processing, collection development, reference, and computer and network operations. The selection process allowed us to involve a large number of Yale staff both from the library and from the computer center and, particularly during the development of the project requirements, to cultivate their knowledge and expertise in imaging systems.

In addition to the involvement of staff, the selection process yielded a highly competitive proposal against which to test the validity of the original master plan. The proposal identified specific hardware, software, services and staff required for the project and it established a cost and budget base for the project. The vendor selection process also afforded us a better understanding of factors affecting the development of the imaging industry, and particularly a more sophisticated view of the feasibility of and risks entailed in Project Open Book.

The Xerox proposal

In the end, the vendor selection process generated for Yale two competitive proposals with Xerox being the clear winner. Three aspects of the winning proposal merit attention here: the technical components of the proposed system, the design principles used in putting it together and its cost elements.

• Technical components

Technically, the system will consist of four main components. First, it will consist of a conversion subsystem for capturing microfilm in digital image form, for structuring the individual image components into a document, and for storing the results on the appropriate medium (magnetic or optical). In the conversion process, scanning, quality control, and structure composition and indexing all require staff intervention. Project Open Book will examine the most efficient way to structure the conversion workflow to minimize errors and maximize speed. The requirements analysis generated considerable discussion of the types of documents, of the varieties of ways to represent the structure for each type, and of the best approaches for indexing the different structures. For subsequent ease of access, conversion staff in Project Open Book will be able to assign each image page in a scanned document its actual page number in addition to its relative number in a sequence of pages.[4] As resources permit, conversion staff in Project Open Book will also make documents accessible through various levels of hierarchical structure files, which will vary according to the type of document being scanned, such as monographs, serials, and manuscripts. Typical structural elements, depending on the document type, will include volumes, table of contents, chapters, notes, folders, registers, and indices.

Second, there will be a storage subsystem for managing the document images and the associated structure files. The storage component will be accessible over the network and will make available the digital image documents in a relatively low resolution form primarily for browsing at 200 dots per inch, and a relatively high resolution form primarily for printing at 600 dots per inch. Optical media will store the images; during use, magnetic disk will provide a data cache for

faster response time. The document files will be stored on magnetic media in a relational database.

Third, the system will support personal computer workstations distributed on the campus network to browse, display and print image documents. The browsing stations will allow users to zoom and rotate images, place bookmarks, and advance to exact pages without necessarily scrolling sequentially through the document. Yale will purchase the Xerox CLASS software product to provide these various user access functions. CLASS is still under development and Xerox plans to release it as part of a future Xerox standard product. As a beta test site, Yale will work closely with Xerox to develop the user interface.

Finally, the imaging system for Project Open Book will provide network access to high quality, Xerox Docutech image printers for reproducing the image documents on paper upon demand. The CLASS software will also support the use of printers attached to the user's workstation primarily for page printing.

- **Design principles**

As part of the technical design of the proposed imaging system, the Yale Library insisted on several principles. We expect to convert documents at the highest possible digital resolution so that the labor of conversion has to be performed only once. We required the use of formal and de facto standards, including the use of the Tagged Image File Format (TIFF) and CCITT Group IV file compression for file storage, and support of TCP/IP ethernet network protocols for document transmission. We insisted on a design that incorporated a client/server approach with documents distributed to user workstations and made accessible through interfaces that are native to the workstations, such as the MAC OS and/or Windows. We also demanded a phased approach to implementation.

The phased approach is designed to give us the ability to invest in the project in relatively small increments, to achieve and evaluate results at each step, and to abandon the project without significant loss if the results in any phase are not worthwhile. The phased approach that Xerox proposed is slightly different from the approach originally outlined in Yale's original master plan.

The first phase will support conversion of up to 100 microfilm documents. Components will include a single conversion station, networking to the Docutech printers, simple document indexing and, for bibliographic information about the documents in image form, terminal access to Orbis, Yale University's online catalog. This phase will provide experience with the flow of work in the conversion process and with image quality, particularly the effects of scanning resolution. It will also allow for the evaluation of user access issues.

The second phase will advance the project into high volume conversion. The components will support expanded structural composition and document indexing, network-accessible storage, and the ability to deploy multiple browsing stations within a library. Evaluation of network access to the image library will begin during this phase. In the third phase, we will continue the high volume conversion activities. The storage system will be expanded to full capacity, and we will aim to deploy multiple browsing stations throughout the campus.

- **Projected cost elements**

The Xerox proposal included cost estimates for the entire project. Pricing for the system implementation assumes the existence of Yale's campus ethernet network and its high-speed image printers. The estimated costs include system

BEST COPY AVAILABLE

integration services, requisite hardware and software, including a license to the CLASS software product, and are estimated at about $1 million over the three phases, which will probably span three to four years.

By the end of the project, annual operating costs for the maintenance of software and equipment are estimated at $60,000. The costs of the labor required for staffing the conversion process, for network and printer usage, and for facilities management of the storage devices and image servers will vary and depend on the level of effort applied to the project and its use.

Factors affecting the development of the imaging industry

In addition to the competitive proposal, another significant result of the vendor selection process that we followed included a better understanding of the factors affecting the development of the imaging industry. Corporate response to current and potential markets for imaging projects varies considerably and key technical issues remain to be resolved.

- **Corporate response**

The Yale Library worked intensively with only two companies and they differed considerably from each other along several dimensions. First, there was a striking contrast in the abilities of each company to identify—and identify with—the needs of the Yale Library as its customer. Each corporate team, for example, brought a highly disciplined approach to the project. In the one case, however, the project methodology drove the team to emphasize technical features that conflicted with or ignored clearly stated functional requirements. In the other case, the disciplined approach emphasized that the customer needs to be satisfied with the results and it thereby helped insure that functional requirements, which the engineering team lost or misunderstood, were eventually recovered or corrected.

Another dimension along which the vendors differed in their corporate response to the imaging market emerged in the resources they had and made available to the Yale project. One company, for example, had two librarians on staff and participating with the project engineers in the analysis of Yale requirements. The other company did not bring specific library experience to the project. The engineers for both companies on the project had extensive experience in the imaging arena. However, the project engineers in one case had apparently worked with a relatively narrow range of products and were unable to reach very successfully for help either inside or outside of the corporate organization. In the other case, the engineers on the project team had considerable experience with a broad range of imaging products and an extensive, responsive network which spanned corporate boundaries and which they tapped repeatedly and effectively for expert help.

Finally, companies of the size and scope of Digital Equipment and Xerox approach new markets from several different angles: from their internal research and development operations, from their sales organizations, and from their custom services divisions. The ways that the companies juggle the interaction of these three different divisions make an enormous difference in their abilities to respond creatively to new market opportunities, such as the Yale and Cornell imaging projects. One company emphasized their sales and custom service division while the research and development operation played almost no visible role; the other emphasized the synergy of its internal research and development operations with its custom service division and de-emphasized the sales component.

- **Key issues**

Despite the significant differences we witnessed in the corporate response to the imaging market, which Project Open Book in part represents, the interaction we had with Digital Equipment and Xerox confirmed for us that the management of complex documents in image form is a general problem crying for solution in many arenas. It is not confined to library preservation, to libraries, or even to academic institutions. Although the market for imaging products is thus potentially broad, our experience suggests that it is nevertheless relatively immature and just emerging. Development of the market will depend on many factors, but in our view it will depend on the successful resolution of several key issues.

Two of these factors deserve special mention, namely the quality of microfilm scanning and the facilities for browsing image documents. First, the quality of microfilm scanning requires technical attention and organizational control. Although it is apparent that the technology for scanning microfilm is readily available in off-the-shelf products, a variety of technical features affect the quality and costs of scanning.

Among the key features, which are subject to considerable variation and which we have to aim carefully to control, is the quality of microfilm, particularly in film produced to preservation standards. As we gain more experience, it may be that we need to modify the standards for preservation microfilming, in part so that gray-scale and color images are not forever lost in the use of high-contrast, black and white filming techniques. Scanning from microfilm will require careful and continued attention to the levels of digital resolution that can be achieved. Other factors that affect the quality of microfilm scanning include the techniques available to the operator for separating two-up images filmed in comic and cine mode and for controlling and correcting image imperfections, as well as the ways for integrating quality control into a digitizing work flow that includes document indexing and storage. Notably, Digital and Xerox differed substantially in their approaches to and assessment of these various factors.

Standards for measuring the quality and costs of scanning, in turn, depend greatly, as we mentioned earlier, on the uses to which image documents will be put, including storage, printing, browsing and, potentially, character recognition. All these uses are points of optimization in a imaging system. But what is especially at issue is the nature of the facilities that an imaging system provides for direct browsing access to the image documents. The facilities currently available for readers to browse image documents are almost certainly the weakest aspect of imaging technology and in the most need of development. Here too there are a variety of factors affecting the usability of complex documents in image form.

In systems like those proposed for Project Open Book, we need the ability to handle in image form the full range of document types, including serials, monographs in single and multiple parts, manuscripts, maps, photographs, and so on. We need conventions for unique document identification, including conventions for designating holding location: is the image document in the ether or where? We have to decide where we will locate the database of record for bibliographic information about image documents. Should we be looking in our on-line catalogs for the bibliographic record or in the imaging system itself? We need to devise conventional techniques for making basic internal structures of documents, such as pagination, accessible to the reader and available for interchange from system to system. And finally, we need to address carefully the physical presentation of image documents on the computer display.

Storage and character recognition, of course, raise other important issues, many of which were articulated in *From Microfilm to Digital Imagery*.[5] At this stage, however, neither issue seems as vexing as those associated with the conversion and browsing issues. Nor do they seem to require as urgently the controlled, hands-on experience, which is essential to provoke the development of a standard set of appropriate technical and procedural solutions, and which Project Open Book is designed to provide.

Conclusion

The Yale Library is now poised to complete the organizational phase of the project. We still need to decide what material we will convert to digital image form. We are likely to draw substantially from our Great Collections of preservation microfilm in either European or American history. We are also actively seeking funding for the first and subsequent phases of the project.

As work proceeds to the next phases in Project Open Book, Yale recognizes the need in the library community to find collaborative ways to address the key issues raised by the use of digital image technology. In particular, it needs to build a technical and organizational infrastructure of equipment, software, networks, and knowledgeable users and staff that spans multiple campuses and facilitates the reliable and cost-effective interchange of image documents.[6] Building on its experience in Project Open Book, the Yale Library expects to contribute substantially to an understanding of the role of digital imagery in the library of the future and to the collaborative efforts needed to insure its effective use.

Notes

Acknowledgments: We are grateful to the Commission on Preservation and Access for its support of the organizational phase of Project Open Book. We found it a pleasure to work with the staffs of Digital Equipment Corporation and the Xerox Corporation and deeply appreciate the interest and support of both companies. At Yale, we gratefully acknowledge the help of the members of the Steering Committee and Project Team and of our colleagues, including Katherine Branch, John Coffey, Dave Gewirtz, Howard Gilbert, Bernie Hayden, Greg Kaisen, John Meerts, John Meickle, Fred Musto, Sandy Peterson, Susanne Roberts, Alan Solomon, Richard Szary, and Alan Watt.

1 Donald J. Waters, *From Microfilm to Digital Imagery. On the feasibility of a project to study the means costs and benefits of converting large quantities of preserved library materials from microfilm to digital images* (Washington, D.C.: The Commission on Preservation and Access, 1991).

2 For further discussion of the limitations of character recognition technology, see *ibid.*, p. 8, note 11.

3 *Ibid.*, p. 9.

4 The distinction between actual and relative page number is critical for document access. Consider a book with 6 pages of unnumbered title material, 28 pages of introductory material numbered in roman sequence, and 250 pages of text numbered in arabic sequence, with several sequences of unnumbered illustrative plates interspersed throughout the text. Now consider the task of finding the plate facing page 157. If the scanned page images are indexed only in a relative order, the reader who requested page 157 would be at least 34 pages away from the desired page (6 unnumbered title pages + 28 introductory pages + any additional unnumbered illustrative plates that appear prior to page 157). If the scanned page images are indexed to the actual page numbers that appear on the pages of the document, a request for page 157 would place the reader immediately adjacent to the desired page.

5 Waters, *op. cit.*, pp. 19-29.

6 Donald J. Waters, "Mission and Goals for a Digital Preservation Consortium," Yale University Library, Department of Library and Administrative Systems, 1992.